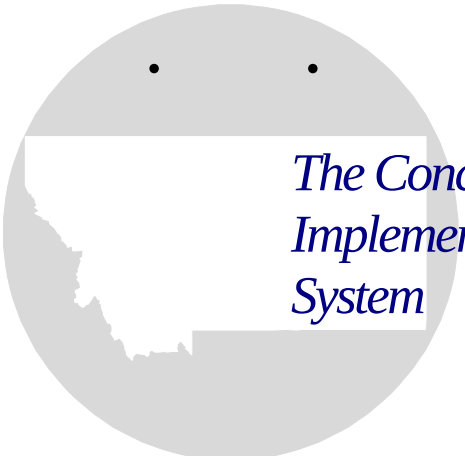
Intelligence Consulting, LLC and Parex Consulting Group, LLC

Concept and Implementation Plan

Montana's Education Reporting System



The Concept of Data Warehousing and Implementation of a Student Record System

June 30, 2004

SUMMARY REPORT

Montana's Education Reporting System (MERS)

Executive Brief

A Need for Leadership and Consolidation of Resources

A data warehouse is a store of data that is designed to facilitate management decision-making. The outcomes of a well-designed data warehouse will go beyond the requirements of NCLB, IDEA and Perkins and will be used to:

- improve the state's ability to collect and report state mandated data;
- allow rapid response to requests and justification for different funding;
- enable local districts to benchmark their own educational programs;
- improve the capability of researchers to study trends for school improvement;
- allow parents and community members to learn more about their schools;
- assist the state in improving compatibility with national data standards;
- include information needed for the ongoing administrative and evaluative activities of the education system;
- provide status and progress toward goals established in the Montana School Report; and
- contribute to indicators that convey knowledge about enduring education issues of our time.

A data warehouse includes not only data but also the policy, procedures, personnel, training, and query tools that make access to the data easier and more relevant to decision makers. The purpose of the data warehouse is to increase the value of the organization's data asset.

A Framework for Change

Change is a process not an event. We can't just will it to happen but need to provide leadership, coordination, and support that will ensure that local efforts can thrive and all educators can participate in the information revolution.

At a point in time the Montana Education Reporting System (MERS) warehouse will become an integral part of the educational community's information gathering process. This will happen when superintendents, boards, legislative aides, analysts, as well as end users, take it for granted that the warehouse is their **single source** of educational information. As a single source, all decisions may lead to a more unified vision of education. To achieve this vision MERS will address three chronic complaints most frequently leveled by school districts:

- Excessive forms: This proposal will eliminate the existing collection of aggregated data collection. This data is often inaccessible and unverifiable, which makes any substantive evaluation of the effectiveness of our system unworkable.
- Inaccurate data: With this proposal student data will be entered as close as possible to the source of that data thus reducing duplicate entry, while increasing data reliability.
- Lack of integration: MERS data will reside in a relationship with other department data sets, broadening the information to legislators, administrators, and district staff in order to make better decisions and improve program policy.

The Concept of Data Warehousing

Building a data warehouse to support decision makers.

The purpose of this paper is to provide Montana Office of Public Instruction (OPI) with a broad understanding of the benefits of a statewide data warehouse and the step-by-step process to initially populate it with single record student data. OPI initiated this process by requesting a design that would address federal legislation (IDEA, NCLB and Perkins), enhance Montana's school accountability system, and assist the OPI in long term policy development. With this paper, it is anticipated that the OPI could design a Request for Proposal (RFP) that would engage a qualified vendor to complete the development of a comprehensive Montana Education Reporting System (MERS) and deploy the system for the educational community by the year 2005.

The term educational community is used throughout this paper to include the group of stakeholders who have a strong ongoing investment in the performance and growth of students. This includes: state legislators, parents, teachers, local and state administrators, businesses, media, researchers, educational organizations, higher education, auditors as well as boards of education and the general public.

How will MERS work?

The MERS provides for the timely presentation of data to a policy maker. It can assist a policy maker in moving from guesswork to confidence in supporting a decision because of the high quality of data, the capability for rapid access, as well as sophisticated data manipulation tools. The MERS is designed to make it easy and efficient to access information in multiple formats, via the Web.

Data are a strategic asset for any agency, but their value becomes truly tangible when it is turned into viable information for everyone to use. The MERS warehouse, with its decision-support tools, can create an opportunity to change the way the educational community in Montana views student achievement, school accreditation and program change.

The Montana OPI began exploring the creation of MERS in 2003. Its purpose is to develop an infrastructure for the educational community to gather school data via the Internet, manage the data in a warehouse that is secure, and make the data accessible to decision makers throughout the educational community. More specifically, Montana is seeking:

- A flexible system that can respond to constantly changing legislative mandates, and has the capacity to accept and safely process files received from the variety of local district student-management packages used throughout the state,
- A manageable, centralized warehouse repository of information to provide accurate student, staff and school infrastructure information necessary for determining school quality and allocating state funds,
- A repository of student data with current, agreed upon definitions that includes technical requirements and business rules for data transformation, to make it accessible, clean and timely to the end user,
- The capability for school districts to have online access for inquiry into their respective student information to help resolve duplicates and respond to reports prior to them being made public,
- Specified uniform management and reporting, as well as access to warehouse information through query and ad hoc tools,
- A cost effective system that will accommodate the existing investments that local districts have made in their own student information systems, and

- An open information system with full protection of confidential data.

It is anticipated that pilot groups and school districts in Montana will provide 'hands-on' experience as the MERS is being constructed and provide valuable, detailed feedback so that their suggestions will improve both the collection process and the reliability of the data being requested.

Data warehousing also enables the creation of more sophisticated Decision Support Systems (DSS) by delivering critical information to be easily accessed by the educational community. How will the community benefit? The answer lies in the realization that information is one of education's largest and most underutilized assets. As the number of staff in the OPI continues to be reduced, the Department continues to amass larger amounts of data; it is information that often holds the key to more effective educational policy. However, the current access to this information is underdeveloped. Large amounts of data exist, but creating a smooth, community-wide access to this data in a timely manner is difficult. It is no surprise that data warehousing is quickly becoming an integral part of the deliberate strategies in the OPI.

What is a Data Warehouse?

Data Warehousing (DW) is relatively new, first emerging around 1992. It is the most significant advance in recent computing and has become one of the most effective decision support tools of the past decade. A review of literature has found that there is still not a clear definition that can be agreed upon. However, within this document, a DW will be defined from two perspectives: depending on which side of the keyboard one is sitting.

> If the person is part of the educational community who is sitting in front of the keyboard then a data warehouse is a tool that allows the user to move from guesswork to validation regarding individual and cohort student achievement, financial management, curriculum design, and teacher quality issues.
>
> If the person is part of the information technology (IT) staff who is sitting behind the screen then a data warehouse is the management of relationships between several data sets, data transformation, security, maintaining data integrity, as well as the timely presentation (push and pull) of data to users in a format of choice.

In either case the goal is the same:

> ***…to reduce data burden and encourage better policy making by maintaining a cost effective method of transferring and accessing accurate and timely educational information among school districts and the educational community.***

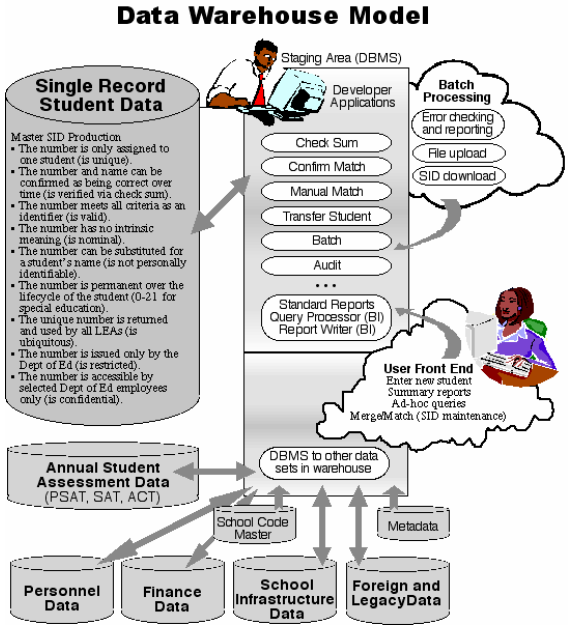
To help understand this concept further, consider this analogy: a data warehouse itself is like a big bucket. Inside this bucket is something similar to Legos®—those small plastic building blocks for children. With Legos®, one can build or rebuild just about anything from an assortment of generic, standardized shapes, such as an action figure, a house, or a spacecraft. The only limit is one's imagination and the number of pieces in the bucket.

Data warehouses work the same way. One can build or rebuild just about any report or table s/he wants from the warehouse's generic pieces of data, and is limited only by the technology tools being used and the data elements stored in the warehouse.

What educational data can be stored in a warehouse?

In the long term the MERS warehouse could contain five sets of data each with escalating security depending on whom in the educational community is seeking access. The five data sets that state Departments of Education initially include are:

- Single student record data (e.g., date of exit and reason, grade, program, unique ID number).
- Student assessment information data (e.g., annual statewide assessment/CRT, GED, Advanced Placement scores, ACT, and PSAT results).
- School infrastructure data (e.g., condition of building, number of classrooms, age of PCs, connectivity, crime/safety, accreditation).
- School personnel data (e.g., degree held, certification, assignment by grade level, salary, FTE, Professional Development).
- District financial data (e.g., comprehensive financial data which includes balance sheet, revenue, district expenditure, and building expenditure).



These five core areas must contain the fundamental elements that are essential to the management, improvement, and development of educational policy at all levels. Over time additional data sets might be added to the DW, such as: university enrollment, employment, census, health and human service data as well as data from older legacy systems. Usually these sets are all joined together relationally by a data set that contains school numbers (i.e., a School Code Master). Appendix A contains a more detailed example of a data warehouse model.

What are the benefits of MERS?

The primary benefit of the MERS is the direct benefit it provides to the educational community. By consolidating and sharing the data, no longer will districts be required to fill out a multitude of paper forms for disparate offices. The consolidation of data entry will reduce errors and misinterpretation. In addition, consolidation often results in fewer collections, uniform formatting, and less annual work by data entry staff which increases the time staff can spend on data analysis for school improvement. Reporting entities will receive a direct benefit by using technology not only to provide the information, but also to view changes and trends that enhance decision making for their agencies.

The MERS warehouse will become the **single source** of consistent data that will describe the educational environment to stakeholders in Montana. Its design is a set of common core data, non-technical fields that all participants can understand.

- The product of the warehouse will be information.
- The power of the data warehouse is in its users who make increasingly important decisions faster than before.
- The outcome becomes collaboration between stakeholders that will enable better policy-making that leads to school improvement.

How does the MERS increase efficiency and lower costs?

The need to manage data as a community asset is not generally recognized. Many agencies are unaware of the benefits of a well-administered data warehouse and the impact that poor data have on policy.

The educational community pays a high price for low-quality data. Policy developers at all levels cannot effectively create appropriate strategies to address the real needs of the education system.

Decision-makers spend needless hours hunting for data, correcting inaccurate data, working around format problems, and scrambling to assemble information across non-relational databases. Education as a whole has had a very poor return on its data investment. Countless hours and funds are spent collecting and organizing data that are often reported only once or are too out dated to have much significance.

The high cost and acceptance of current data systems promote a sense of futility with the way things are. Because of this, our educational community has come to accept the level of low-quality data as normal or typical behavior of a bureaucracy.

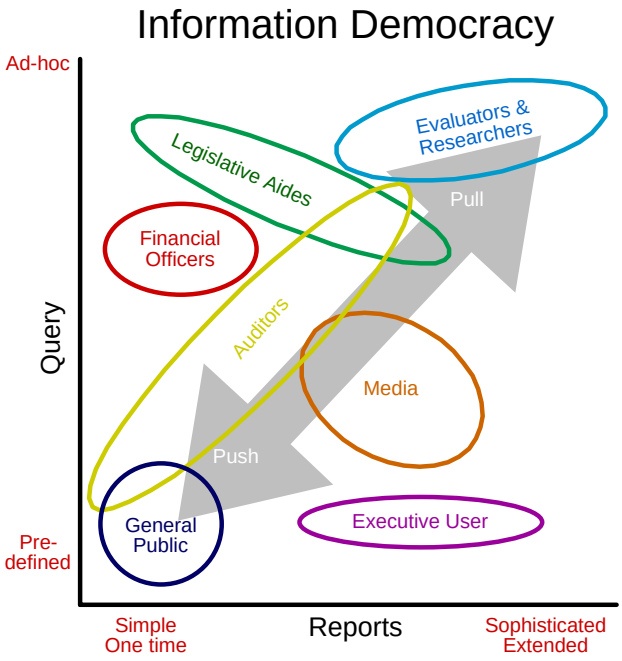
Educators can no longer afford the high price of low-quality data. In this age, a reliable data warehouse means the difference between poor and effective policy. Even when the value of good data is well documented, the need to manage it as a community asset is not generally recognized.

The MERS warehouse is cost effective because it breaks this cycle of low-quality data, provides an integrated environment of usable data, and makes the best use of current technology. Well over half of the cost of the current data system is attributable to low-quality data management found in the existing environment. The MERS warehouse will be cost effective because it will integrate data from disparate databases, eliminate redundant input, and consolidate data gathering techniques.

Simply stated, the higher the degree of partitioning and the longer data are kept in separate silos and redundant data are kept by different agencies, costs will continue to increase and school improvement and student performance will continue to be less manageable.

Creating an Information Democracy

Data warehouses promote the concept of what can be termed: *information democracy*. This means that everyone (with appropriate security) can access the data in the MERS warehouse (i.e., it is an open system). Different stakeholders in the educational community need different access to the data. For example, members of the general public often need summary or 'canned reports' that are regularly pushed to them from the DW, whereas researchers or legislative aides may need online access to conduct a series of 'ad hoc' queries.



An *information democracy* can best be visualized using the chart on the left. It depends upon the number of questions (called queries) that can be asked by the nature of the reports that groups want to produce (i.e., how many and how often). This continuum must be available to any group when needed.

In such a democracy, less experienced data users must have reports regularly pushed to them via the Internet and those groups that have the capability for greater analysis may have permission to pull more selective data at any time from the DW.

However, the educational community must keep in mind that a data warehouse in and of itself will not improve the performance of school's, school improvement is derived from careful analysis of the data in the warehouse along with the decisions that it supports.

Simply put, an information democracy cannot be approached with the same IT management principles that are currently in practice. The MERS warehouse must be a centralized repository of data structured for easy and direct access by the entire educational community. Today, data users expect to participate in an on-line, interactive dialogue–instead of receiving untimely generic paper reports that answer only static questions.

Implementation of a Single Record Student Database

The key assumptions about collecting student data are based on simple principles designed to improve the quality of what currently exists and based on the belief that:

- The student data must be collected on a regular and timely basis. There is no efficiency in gathering a student data element that reflects a one-time need or an unusual bit of information.
- The data must be reliable. Student data should reflect consistent measurement of an entity from one time to another.
- The data must be valid. A student data element must represent a logical and meaningful portrayal of the subject without distortion.
- The data must be consistently defined. The data elements should have a standardized definition so that each can be collected from districts in a systematic manner.

The primary reason for improving the quality of our educational data is to improve student performance and overall public satisfaction with our educational system by increasing the efficiency and effectiveness of our educational policy. We can no longer judge the educational system by looking only at output measures. We must begin to evaluate its outcomes as well. This means that we must use student data to go beyond input, process and output issues, and address effectiveness issues (see Data-Based Decision Model in appendix B.)

There can be as many as twelve steps to build a state level single record student management system. The foundation of MERS begins with the creation of a comprehensive policy with adequate funding that commits the organization to a strategic path and ends with the ability of a user to do extensive 'data mining' that supports good educational practice. Many of these twelve steps can be implemented concurrently; a few steps will cost much more than others.

Step 1: Develop Policy

Step #1 involves a commitment to bring the Montana educational data management system into the 21st century. Establishing the correct foundation to collect single student records involves the creation of a variety of policies over several years. Three of which are discussed here include: a concurrence among the educational community to collect individual student data (vs. aggregate), as well as security and funding issues.

To illustrate the need for policy, first consider the collection of individual student data (also called discrete or primitive data.) If a key element of a data warehouse is to deliver information, then it is important that the DW model only include primitive data, or at least the lowest level of detail as possible. For example, it is more useful to have a person's date of birth rather than his/her age. It is more useful to have the gender and grade of a single student than an aggregate table that includes the gender of all students by grade.

> The first reason for including primitive data is that aggregate data (data that is already summed-up) limits the user's ability to redefine it and create different tables. For instance, if the database only contains annual FTEs by grade level, then a user could not find out whether the FTE in a building is lower or higher in the beginning or the end of the school year, in a larger vs. a small district. This is because data that is already derived (totaled) cannot be disaggregated (e.g., once a chocolate cake is baked the ingredients cannot be taken apart to make brownies.)

A second reason for including primitive data is called drill down. Drill down is a capability found in all business tools where the user can access the details behind a derived field. This data mining capability enables a 'peeling the onion' advantage in obtaining information, where any limit to greater detail in the data set would frustrate the user.

The third reason for including primitive data is that it is much more difficult to model the unlimited configurations of aggregated data. The quantity of aggregated data elements is, for all practical purposes, infinite when compared to the configurations of primitive data. The choice to limit a collection system to aggregate data means limiting users to the queries and analyses found only on static 'forms' or on a standard format gathered electronically (e.g., think of aggregate data as a photo – one shot in time – and primitive data as a movie – has a beginning and an ending.) Such a decision causes serious problems when reauthorized legislation mandates new data collection forms or different reports.

A corollary to collecting primitive data is the capture of data at the time of initial enrollment of a student. This is critical to the individual that will enter the student data (usually the building secretary) into the district's Student Information System (SIS). Uniform collection of initial student data will increase accuracy and reduce the burden on districts that may have to enter or reenter the missing information at a later time.

A second policy issue is security. Every organization has data that are proprietary or sensitive, and a data policy must develop a security system to protect and distribute the data. These policies must indicate that:

- School districts or other primary sources of the student data are the originators and owners of that data. The OPI functions as the custodian of the data. Data obtained under a request are considered a loan and may not be sold or rented and that commercial use of data is prohibited.
- In order to protect the data in its custody the OPI will approve a policy that ensures all student data in the DW are securely maintained with safeguards on all personally identifiable information.
- Confidential data on an individual shall not be created, collected, stored, used, maintained, or disseminated by the OPI in violation of any federal or state law.
- If the OPI does enter into a contract with any private person or third party to perform any student functions, that agreement shall require that the data be protected in the same fashion stated in the OPI policy.

Furthermore, in an information democracy, organizational data must be shared, but departments have difficulty sharing. Policy must be prepared to address disputes regarding access, processing restrictions, and other matters (i.e., a forum for resolving conflicts that apply to the entire organization.)

A third example of the need for data policy concerns funding and return on investment.

The typical approach by many State Education Agencies (SEA) is relatively unsophisticated. The system is often composed of a number of spreadsheets, databases, and paper reports (many moved online for data entry) that are loosely connected through various interfaces or patches. The high cost to this approach impacts both the state education agency as well its local districts.

Costs to the SEA include the ongoing programming and maintenance requirements as well as staff time to manage the capture, manipulation, and storage of the data. As the system expands hardware is often not scaleable or worse performs inadequately. The SEA will compensate for this by purchasing more servers and hiring additional staff to ensure timely collection and reporting. As the staff grows, more time is expended on input activities (i.e., redesigning the collection forms, programming, integrating systems, and changing data architectures) then on rethinking its overall reconstruction to a more dynamic system (which are often minimized until its too late).

Costs to a local district are incurred primarily through the horizontal and vertical movement of the data. A single data element usually flows through approximately 10-12 transfer points, from data retrieval, to data entry, cleansing, and transformation to the state level (a process that is often repeated when errors are discovered and retransmittal is necessary). At each transfer point the local district incurs a cost (e.g., recoding, system interface designs, manual reentry of aggregate data, upgrading and maintaining applications, etc.).

It is common with this approach for state agencies to end up promoting extensive human resources at the front of the system (collection, data entry, cleansing, transmittal) and limit the resources needed at the back of the system (analysis, reports, ad hoc queries, decision support training, etc.). The problem with fewer resources at the output end is that it reduces the pace with which districts drive academic change or sustain improvements. SEAs as well as local school districts that lack research and decision support staff or fail to train educators to use tools to access the DW on their own, often fail to uncover critical opportunities that boost student achievement and continue school reform momentum.

The cost to define and implement a data warehouse as described in this document can be considerable – often a 2 to 3 year phase in - when all goes well. However, one must assess up front whether the Department has the internal expertise and competencies to redesign their system and if it can commit the financial and human resources over time to manage, maintain, and improve a data warehouse. Costs are based on the degree of customization, integration with local district management systems, time sensitivity to implementation, and the complexity of the state data sets being sought.

An agencies data are as much a resource as are its building, technology, or financial assets. Data are time-consuming and expensive to acquire, and must have utility beyond operations. Information derived from data can be used to assess the quality of staff, student performance, or effectiveness of programs over time. Unfortunately, as long as the data are locked in data silos (often called data tombs), their utility is limited.

Policy must also increase the agencies return on its data investment. What are the issues? Are we getting what we should be getting from our data resource? If so, can we get more? Sometimes setting policy also means investigating new techniques for storing, processing, or presenting data.

Data warehouses are designed to be the archives, and saved for a very long period. Storing data for more than twenty years is very common in an education warehousing system. The cost of maintaining the data once it is loaded in the data warehouse is very minimal. More significant costs are incurred in data transfer, error checking, creating the unique student ID, deduplication of students, audit, and purchasing warehouse-reporting tools.

The return on investment by using a data warehouse is directly related to the data users' ability to perform three tasks: access the data, analyze the data, and produce reports in a more timely fashion.

Recommendations and Cost
Step #1 is a low cost item provided that local districts are already using student management packages for scheduling, transportation, teacher grading, due process, or integrated reporting needs, etc. In a recent OPI survey it was reported that (insert results of survey, if available) % of the local districts in Montana are still using paper based management systems. There are several approaches that can be used to bring all districts in Montana to a level of student data collection needed for MERS. Alternatives include: statewide purchase of a single SIS, reimbursement to districts to purchase an SIS of choice, or building a Microsoft Access database for local districts to enter only the minimum student data required by OPI.

Step 2: The Data Resource Administrator and the Data Base Administrator

Step #2 involves assigning personnel to oversee and manage both the DW and student record system. This involves assigning a Data Resource Administrator (DRA) and a Data Base Administrator (DBA).

The four primary responsibilities of the DRA are maintaining the Meta data, producing reports, ensuring that appropriate statewide training is ongoing and managing all stakeholder groups in the educational community. The three primary responsibilities of the DBA are managing the performance of MERS, building (or supervising) application development, and maintaining data integrity.

To understand the need for data administration, consider the analogy of a high school library. The typical high school library contains hundreds of books, journals, magazines, reports, and so forth, but they offer no utility while they are on the bookshelves. To be useful, they must be made available to students who have an interest in and need for them.

Clearly, the library must have some means of describing its collection so that potential users can determine what is available. At fist glance, this might seem like a trivial problem. You might say, “Well, build a card catalog.” But much work must be done to be able to do that. How should the library’s works be identified? How should they be described? How can we accommodate different ways of identifying works (ISBN, Dewey decimal system, government report number)?

Furthermore, suppose the district is so large that it has several high schools. In this case, how are the library collections to be managed as a single resource? Furthermore, middle schools maintain their own libraries. Are these to be made part of the district library system? Many teachers have extensive personal libraries. Should they be part of the system?

Administering a warehouse that holds several collections can fail for a number of reasons. When it does, they are usually traced to inconsistently defined fields, inability to relate fields from different data sets, unacceptable query performance, lack of timeliness to the data, and perhaps the most common reason—the absence of a good Data Resource Administrator.

Moreover, to be successful, the entire community must make a commitment to student data management as a discipline. The DRA must provide enough information and presentations to key organizations (superintendent associations, teacher unions, legislative subcommittees, parent groups) to ensure they fully understand the MERS and must organize the data management meetings (with key program and technical staff from local districts) and the steering committee (with special interest groups and statewide educational organizations) to ensure that all the operational aspects that users need are being satisfied.

The DRA will develop and enforce standards, and sustain an environment for the community that encourages strategic planning and better interpretation of student data. The position facilitates policy-making and establishes priorities. The DRA has the authority to:

- approve modifications to the model and allocate access to the database;
- control codes, file format, record layout structure and data integrity;
- work with the stakeholders on data definitions, edits and data transmission;
- support data users with a correct understanding of its contents; and
- design special studies that target educational issues that are too great a data burden on districts to gather through annual statewide collection methods.

The DRA must work with the DBA to connect student records with different data sets in order to provide more complex information to the educational community. The DRA does this by identifying commonalities among the data sets before they are brought into the warehouse. These commonalities are called 'relations'. In a relational database, the rows of a table represent records (i.e., a student) and the columns represent attributes (i.e., a date of birth or gender).

OPI must assign a Data Base Administrator (DBA) to work with the DRA to optimize the use of student data to its full capacity. The DBA has responsibility for the technical implementation of the database environment, the day-to-day operations of the database and the policies governing its everyday use, including: security systems, data integrity, monitoring and measuring system performance, reporting problems, establishing audit trails, ensuring maximum online availability.

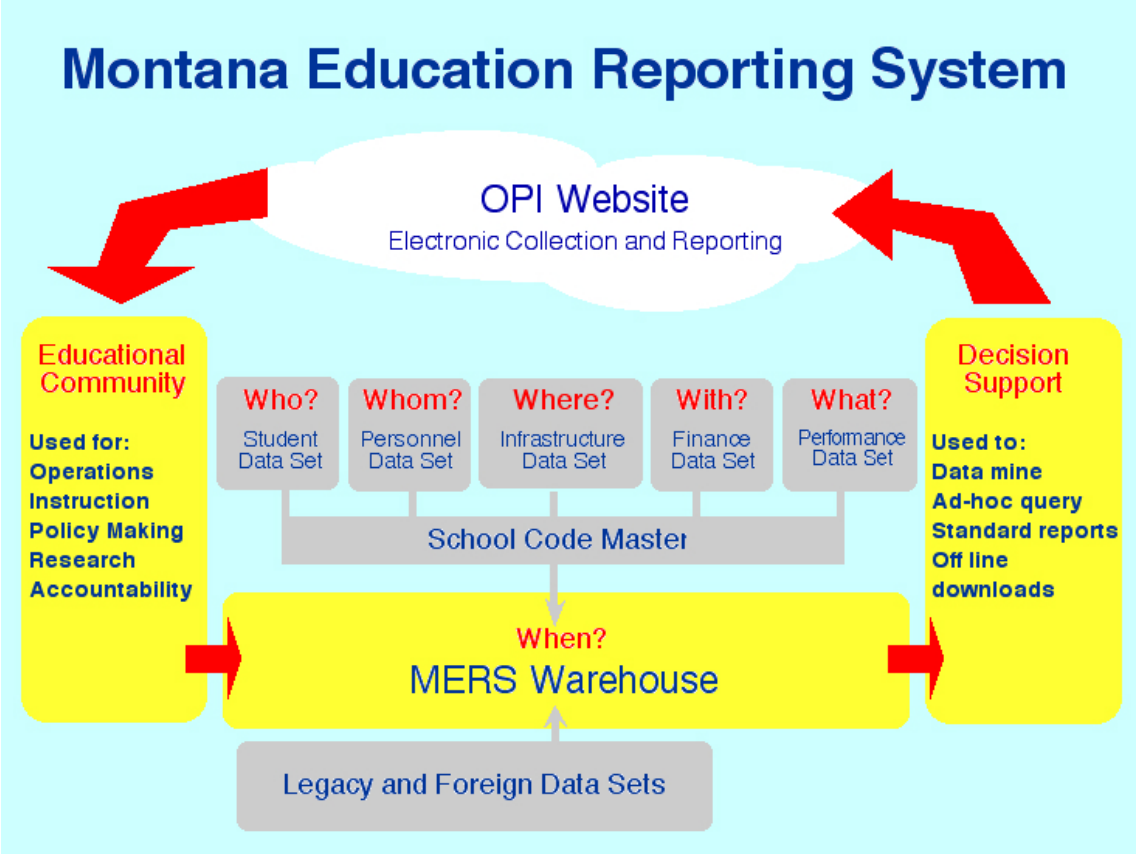
In this respect the DBA continually assesses the performance (speed and capacity) of the database for maximum effectiveness. The DBA must also recommend new tools, technology, and products for the development or enhancement of various applications in both accessing and processing the information on the student database.

Recommendations and Cost
Step #2 can be low cost if existing staff can be reassigned. It may be possible to start a DRA and a DBA half time on the MERS and increase their time as the warehouse grows. In 2004, the cost of two part time staff would be approximately $50,000; two full time data administrators would be $100,000.

Step 3: Define the Data

In step #3 the student data elements that need to be collected must be defined.

The data in the warehouse is often grouped into common sets (i.e. performance, personnel, student, fiscal, infrastructure) and provided to the user in a read-only environment. To make sense of these data sets, a Meta data dictionary (the data about the data) is an absolute necessity. Meta data give a description of and meaning to the fields and values of each record, so that the user can understand what data to manipulate and where to locate it in a warehouse. The data sets in the warehouse are aligned along data "subjects" rather than by functions. For example information about a categorical student (i.e., special education, Title I, Career Tech Ed, etc.) may be grouped together, even though these data would used across multiple reports (i.e., AYP, financial distribution, state report card, etc.) With the right elements in each data set, a series of different databases that originally served multiple functions can become a full relational model. An example of how Meta data is incorporated into the MERS warehouse can be seen below.



Without Meta data, locating information contained in the student record system becomes a daunting task, akin to searching through a bucket full of Legos® for the right piece without any instructions on what the user is building. Meta data not only describe the contents of the system but also provide the user with information useful in judging the quality of the content. It might also contain the date of the last revision or how it was derived from its source.

A warehouse cannot be constructed or managed without an active Meta data catalog for each data set. Because student data is meant for the entire community, and the amount of data can be so vast, data users must also have online information to find what is new, as well as timely notices regarding its refresh schedule. Without an online directory, only hit-or-miss efforts will select the right fields to use (see the Meta Data Manual in appendix C). The MERS will contain data dictionaries, both as a stand-alone online application and integrated into the DW menu.

Wherever possible student data elements must be matched to the NCES Standards as defined in the *National Center for Education Statistics Student Data Handbook for Elementary, Secondary, and Early Childhood Education: 2000 Edition and 2001 Update*. However, because the NCES handbook goes beyond those fields and elements commonly found in most student management systems used by school districts only those used for federal and state reporting, compliance or audit activities are likely to be selected for capture.

One immediate decision that will have to be made regarding student data is the use of race/ethnicity codes. The Office of Management and Budget (OMB) has made available the *Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity* to help in the collection and reporting process. In order to be responsive to federal reporting requirements, OPI will need to decide if the NCES codes will be changed to the Office

of Civil Rights (OCR) codes for race/ethnicity. The OCR codes allow reporting on more than one race/ethnic category, so immediate attention in the development of the Meta data must be given to the question of how multiple race/ethnicity data would be tabulated. Currently, the NCES handbook has five discrete categories (code 150 page 39) and the format proposed by OMB includes any multiples of six categories. For a full discussion on how to report student race/ethnicity using the '97 regs see appendix D.

With Meta data, the educational community will not be confronted with precisely the problem that the MERS warehouse was intended to solve—different answers to the same question with the resulting lack of confidence in the results. With proper maintenance of Meta data history through time stamping, the changes affecting the data in the warehouse will not be lost, and the passage of time will not erode the value of historical data.

Recommendations and Cost
Step #3 is an item that was produced and is being maintained by the OPI staff. OPI completed the online Meta Data Manual at a cost of $12,000, under a grant issued in 2003.

Step 4: Building Security and Maintaining Confidentiality

Step #4 consists of maintaining confidentiality in the student record keeping system.

Confidentiality consists of how personally identifiable information that is collected by an authorized agency is protected. The Family Educational Rights and Privacy Act (34 CFR Part 99 amended P.L. 103-382) referred to as FERPA guards the confidentiality and access to designated records.

A brief explanation on how OPI proposes to establish security and access procedures that comply with FERPA are described below, a more detailed draft can be found in appendix E. To ensure protection:

- The OPI must identify an administrator of the student data who is the designated authority to establish and maintain a system of data protection in accordance with Family Educational Rights and Privacy Act (FERPA), and,
- Each local school district must have on file their district FERPA policy clearly identifying what student data at the local level is "directory information."

Student data must be restricted using two measures: security and data protection.

1) **Security** includes the measures put in place by the OPI to ensure that records are not lost, stolen, vandalized, illegally accessed, or otherwise rendered useless. Since the data are stored on computers, it is essential that there be a high level of protection that provides integrity and availability commensurate with the level of risk and magnitude of harm. Typical security includes automatic encryption and secure socket layer during data transmissions. Certainly, copies of data, and all reports must be maintained in a secure environment to prevent unauthorized access. Secure environments include any electronic media, personal computer, server, network, or warehouse on which the data reside as described in the National Institute of Standards and Technology (NIST) *Federal Information Technology Security Assessment Framework* (November 29, 2000). In addition to the 128 bit encryption (HTTPS protocol), OPI will institute the following 'A' levels of security:

- Assurance/Identification: so that the OPI is confident who is contacting the system,
- Authentication: so that the OPI is confident of the source of access,
- Authorization: granting escalating access rights to student data,

- Access control: methods of establishing user profiling on the DW,
- Automatic log-off: set to an agreed upon number of minutes,
- Administration: recommend ongoing security procedures to the DW, and
- Auditing: provide information on monitoring (authorized and unauthorized) and detection techniques.

An often-expressed concern of educators when considering the wider access to student data is that it provides the potential for greater misuse. Misuse and misinterpretation already occur under existing systems and will be no different when data is more accessible. Creating a totally secure environment that stores student data is unrealistic, but instituting a *trusted system* (i.e., one that while not perfect is trustworthy) is acceptable. The reason for this limitation is that any agency simply cannot counter all risk. When could accepting risk make sense? It is theoretically possible that an asteroid could smash into the earth and land, of all places, on the OPI data warehouse. The risk is real, albeit small, and can be estimated as such. Should the OPI build a concrete vault two miles beneath the surface to store student files, or should one accept the risk of an asteroid strike and figure that the student database will be the last of the worries should the event actually occur? Security assessments and common sense will probably dictate that one can safely afford to accept the risk of an asteroid strike. Maintaining a secure student database does not have to counter every conceivable risk, only those that make sense.

2) Data protection measures include **restricted access by roles** to the student data. Restricted access significantly limits who can view the data and for what purposes. While any number of access permissions can be given to any number or groups of individuals (i.e., users can be assigned one or more access by roles) seeking access to the student data, usually access is limited to five levels. Level 1 is the highest level of access. All access levels are assigned in a way that maximizes public usage without risking disclosure of personally identifiable information.

Level 1 access allows the OPI to read and write to all the student records and fields in the database. This level is only permitted to a minimal number of authorized staff whose members operate or manage the data warehouse or are responsible for maintaining the accuracy, security, and audit corrections in the performance of their duties. Authorization by the security administrator would include individuals such as: the OPI Data Resource Administrator, the state's Warehouse Manager, a State Auditor, or the Security Officer. These staff would all be subject to Montana Acceptable Use Policies.

Level 2 places limits on access to individual student records but not fields. Specifically, superintendents (or designee) of local and county school districts will have read-and-resubmit access to records of their own students. Superintendents will not have access to student records outside of their districts. The purpose is to allow region, county, district, or program levels to verify their own data. Local and county districts are the originators of the data and are allowed read-and-resubmit privileges in the data warehouse to change the records they have already submitted.

Level 3 places limits on fields. The most sensitive fields are excluded from access at this level, such as a student's name or disability. At this level, the unique student ID number is put into practice. This read-only access level is given to a limited number of OPI staff, primarily for the purposes of audits, operations, accreditation, and reporting to state and federal government agencies. Other professionals outside of the OPI may obtain authorization at this level for research purposes. Authorization at this level is for the sole purpose of increasing the existing body of knowledge based on **legitimate educational interest** and not on commercial purposes. **Legitimate educational interest** is defined as an endeavor meant to further the understanding of

educational practices, method, and/or theory that are expected to be analyzed through formal, accepted research practice and the results of which will be disseminated in such a manner as to benefit the educational community and/or public in general. Researchers must submit a restricted access form that explains the purpose of the research and how the researchers will ensure data confidentiality and security.

The fields that are available at level 3 must be disseminated by the OPI to assure that local district FERPA policies and state policy do not conflict when it comes to providing "directory information" to the public.

At levels 4 and 5 data are either presented or provided in an aggregate format. The OPI may produce summary reports from individual student data that relate to groups of students, rather than to single individuals. While it may seem that the use of anonymous aggregated data poses little threat to confidentiality or privacy, there are some cases where populations may include only a few individuals. **Statistical disclosure** is the risk that arises when a population is so narrowly defined that tabulations are apt to produce a cell small enough to permit the identification of a single individual. In such cases, the OPI must apply statistical cutoffs. For instance, if a search were done for the math scores of all Native Americans, and this search revealed two students in a particular building, and then there would be some certainty that information about an individual would be disclosed. A possibility of inadvertently reporting personally identifiable information on these students is eliminated in both level 4 and 5.

Therefore, OPI will block any aggregate results with a statistical cutoff in which **six or fewer students might be disclosed.** Six records is also the minimum number of cases needed to run statistical formulas confidently.

Level 4 applies to state government agencies other than the OPI, as well as state legislators, legislative aides, and the executive branch. Individual student data **will not** be accessed by anyone at this read-only level. However, it is usually possible to perform limited data mining of the core data sets to produce aggregate reports containing averages or totals that relate to groups of students for NCLB. For instance, it will be possible for a legislator to search for fourth grade math scores for the students in his/her legislative district by age, gender, or race/ethnic code. However, s/he could not see the math scores of any individual student or the class performance of an individual teacher.

Level 5 allows read-only access to the general public, including educational associations, media, real estate agents, businesses, interest groups, etc., to view standard reports and data tables that are produced and published in aggregated formats on the Web. Levels 4 and 5 have been separated to facilitate future changes based on locations of legislator aides or policymakers (who are usually inside the State firewall) relative to the general public (who are outside the State firewall).

Recommendations and Cost
Step #4 is a low cost item that involves distribution of the OPI procedures used to secure and maintain the student data in compliance with FERPA. Dissemination of material can occur at the time that MERS is rolled out, along with the training (see step #11) associated with it.

Steps 5-7 involve more technical material
In the following steps #5 through #7 the DRA must work closely with the DBA in the design of

the student Web applications for cleaning and transferring student data, including an audit application (see the Audit Model in appendix F) that will log all access to a student record that ensure that the transfer of student records from district to district includes: (1) The nature and substance of the information released; (2) The name and signature of the official custodian releasing the data; (3) The name of the person requesting the data, the capacity in which such a request has been made, and the purpose of such request; and (4) The date of the release.

While decision-making within the educational community has not traditionally been 'data based,' 76% of the states have or will hire contractors or purchase software to help build their student systems for their state (NCES Forum Survey, 2002). Unfortunately, contractors do not fully understand schools and education. Understanding production rates and financial systems is a great contrast to learning systems used by educators'. However, frequently education agencies do not have the technical expertise or understanding of warehouse technologies as well as large scale student databases.

Recommendations and Cost
Steps #5 through #7 will cost approximately $1.2 million dollars at a price of $8 per student. In many states these funds were used for both internal staff and external contractors. A combination of inside and outside expertise works well because parts of the student database must be brought online quickly; this is what contractors do best. However, internal staff are needed to learn the system as it is being built and contribute to the design during construction so that they can maintain it once acceptance testing has been completed. These steps will also need a high degree of oversight of the contractor(s) by technical staff at OPI.

Step 5: Creating the Unique TID

Step #5 involves the creation of the unique Test Identification Number (TID). This Test ID number is used during the annual statewide assessment to assure that the results of the assessment are assigned to the correct student so that longitudinal performance can be accurately tracked. The TID must be random, computer generated, contain no embedded meaning, and become permanent after being checked for duplicates In addition, after the TID has been appended to each individual record, the student's name must be removed from access and secured at the highest level. To protect personally identifiable information the remaining portion of each individual record now contains only the number. However, the TID must also be controlled with the use of a check sum to assure reliability such that entry of an invalid TID (when passed back to the district) into their local Student Information Systems (SIS) could not occur.

Additionally, the use of a random TID will make determination of a student associated with that TID much more difficult while still satisfying FERPA requirements. In many vendor applications where Student Information Systems (SIS) use a sequential number and the number for a particular student becomes known, it is possible to determine numbers for other students in that school or district since these datasets will frequently be received and processed in alpha order by last name.

One of the outcomes of the MERS project is to develop an TID with the following characteristics:

- The number is only assigned to one student (is unique).
- The number and name can be confirmed as being correct (is verified via check sum).
- The number meets all criteria as an identifier (is valid).
- The number has no intrinsic meaning (is nominal).
- The number can be substituted for a student's name (is not personally identifiable).
- The number is permanent over the lifecycle of the student (0-21 for special education).

- The unique number is returned and used by all local education agencies (is ubiquitous).
- The number is issued only by the OPI (is restricted).
- The number is accessible by selected OPI employees only (is confidential).

In most states the TID is a ten-digit number, where the tenth digit is a check sum (see Check Sum in appendix G). The check sum will be based upon the students first and last name. The reason that a check sum is employed is to assure that the TID and the student name remain consistent and accurate as the student matriculates the K-12 education system. This helps to maintain the integrity of the data. Data entry errors relating to the unique TID often occur at any one of four points:

- After the TID is issued by the OPI, passed back to an Local Education Agency (LEA) and entered into their local SIS,
- When a student transfers from one LEA to another and the TID is reentered into a different SIS in the new LEA,
- When the same student moves from level to level (elementary to middle to high school) within a district and the student data is updated, modified or transformed, or
- When the student data is updated, modified, or transformed.

Applying a check sum digit will ensure the TID remains valid over time (for longitudinal purposes) and that the students last name or first name has not been accidentally changed at any of the above three points.

The need to pre-identify test response forms so that student assessment (i.e., both those tested and not tested) can be tracked longitudinally as required by NCLB is one of the first uses of the unique TID. One of the problems in pre-slugging test response forms is the lead-time needed to print the TID on the forms and the time the test is actually administered. During that window many students transfer between districts or new students may move in from out of state. In such cases a process is needed to identify whether a student has the correct response form and if not issue a new TID to put on the form.

It is for this reason that the official student number itself must always be maintained in a master database. All applications needed to process student records or confirm student ID numbers are kept in a separate staging area. Frequently, testing companies issue their own IDs for students that must be verified against the master ID. Student ID numbers issued by testing companies (e.g., annual performance/CRT tests, PSAT, SAT, ACT, etc.) must be maintained in a different area so that they will not contaminate the master ID. This approach will improve the reliability of the database and help OPI in the construction of the Adequate Yearly Progress (AYP) reports as well as the State Report Card.

Above all, a student database must have integrity (i.e., the data must be logically consistent.) Poor data integrity often results when data are duplicated. For example, if a student changes his or her name or address, then all the files containing that data must be updated, but the danger is that all of the files might not be updated, causing discrepancies among them. Data integrity problems are serious. When data from two agencies is combined merging becomes a serious problem because:

- the agencies may use different identifiers,
- the only common fields may be types with different field lengths,
- common fields may not always be present, and
- the common fields may contain data entry errors.

If data items differ, they will produce inconsistent results. When a report from one data set disagrees with a report from the same data set, who will be able to tell which one is correct? The

results create a lack of trust by the educational community, raise concerns about the quality of decisions that administrators make based on data and the credibility as well as the value of the overall DW comes into question.

Creating the unique TID is easy; resolving duplicate students in the MERS is more troublesome.

Step 6: Checking for Errors in Student Files

Step #6 involves the creation of an extraction, transformation and loading application (ETL). ETL is used to batch process large sets of student data at the initial assignment of the TID and later from large school districts as well as the annual enrollment of Montana's 11,000 kindergarten children. This application, based on clearly defined key fields, would consist of four components:

- Secure logon, including email address to return the status of the processed files, and an error report in plain English that explains what needs to be corrected in the student record.
- An error site that is operational 7/24, where districts can upload and check their student records anytime prior to the official submission date stated by the OPI,
- Error messaging. Feedback regarding erroneous data, reported as Fatal Errors or Warnings, including summary reports: validation (120=120) and verification (numbers are within acceptable ranges),
- An application that permits the district to correct student records that are invalid and need modification prior to resubmitting them to the OPI.

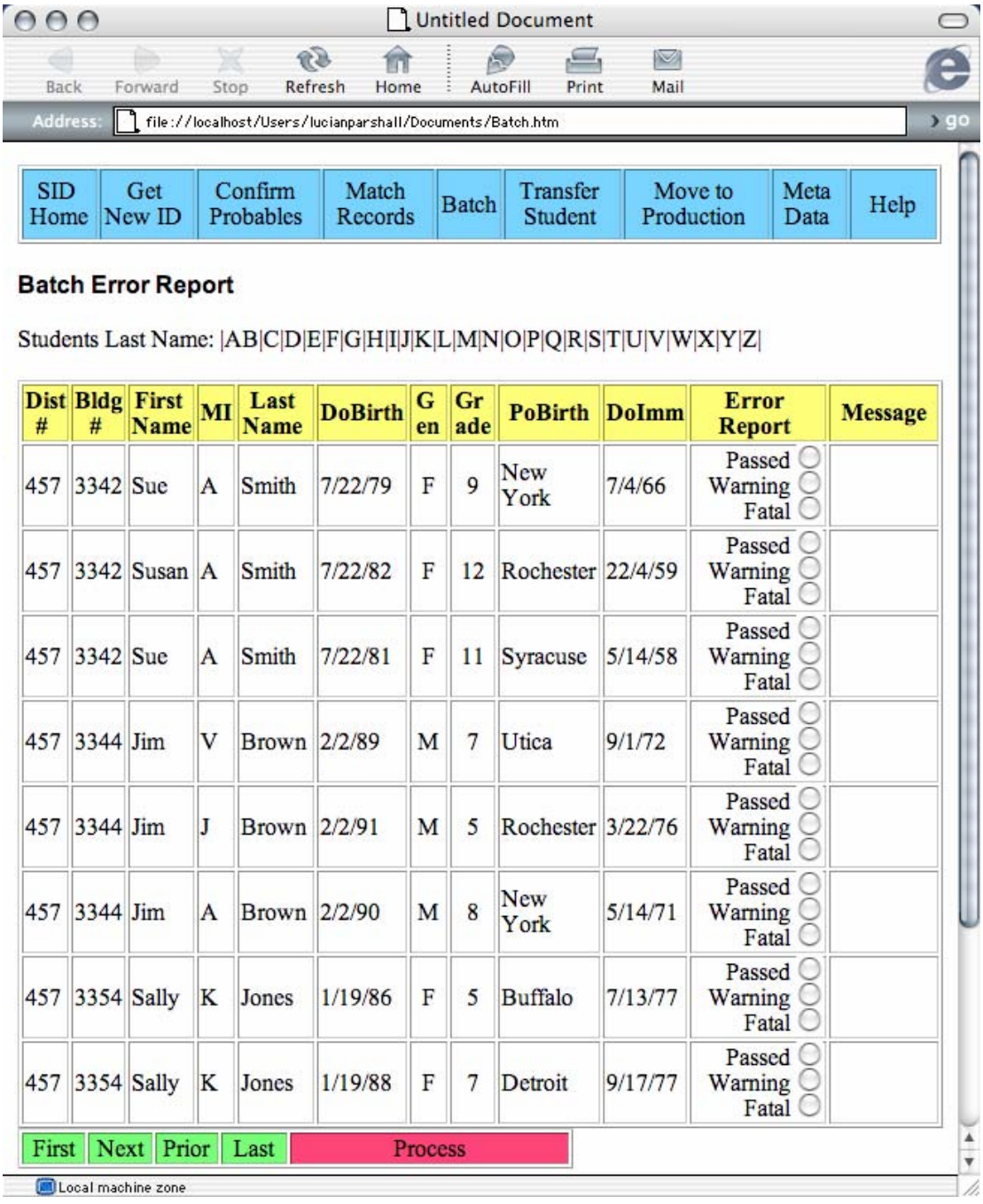
One of the initial online applications that must be developed is one that will transfer student files from the SIS that schools use to manage their student records and transmit these files to the warehouse staging area for processing. This batch application (see example 1.1 below) would perform an edit check routine on the files that generates a sub-folder containing all the good student records (i.e., those records that do not contain minor errors or 'fatal' errors). Suspect records (i.e., records that may contain a duplicate student) as well as all bad records (i.e., rejected records) would be placed into a sub-folder, separate from the good records. This will allow data managers in local districts to easily identify records that need correction. The bad records will contain feedback to the user about why the file was rejected including:

- Error Summary. A report of the field numbers listed for those fields that contain errors.
- Error Detail. A report that lists fatal errors and warning messages for fields collected during other submission periods. Error categories are defined as: Fatal, Warnings, and System Errors.

  Fatal Error - An error that causes the record to be rejected during the error check program and not included in any subsequent processing by the districts SIS system.
  Warning - An informational message that one or more fields of a given record meet the edit requirements but may contain inconsistencies, or discrepancies. The data are often allowed to pass the error check routine and move the record onto the next level and may receive a temporary ID.
  System Errors - Unexpected problem (bugs, missing element, blank data, etc.) within the application (file cannot be processed further).

The same Web site location can also be used to upload the student files to the MERS once they have been cleansed and download new TID numbers after they are issued by OPI. Users would first complete an individual security agreement prior to gaining access.

Batch application example 1.1



Step 7: Resolving Duplicate Students

Once the unique TID is created step #7 begins the process of resolution of potential duplicates.

Creating a unique TID requires the selection of specific fields that will not change as the student transfers from district to district. These fields usually include: last name, first name, date of birth, gender and other unique fields that may be gathered such as: SSN, place of birth, date of immunization, or mother's maiden name, etc. Due to the 1997 OCR directive and the likelihood that race/ethnicity will be a field that is self-selected, race/ethnicity will no longer be a stable data

element for use in de-duplication.

Another straightforward dilemma is the resolution of student IDs who are identical twins, however, duplicate resolution becomes highly complex when the state produces over 150,000 student numbers for its 459 districts. It is for these reasons that local district student ID numbers generated by the SIS become an important part of duplicate resolution process.

Two issues arise that must be addressed when the TIDs are issued for the first time:

1. Including more fields than needed in the creation of the TID increases the number of potential duplicate students to resolve when data in those fields are **inaccurate**.
2. Including more fields than needed in the creation of the TID also decreases the number of potential duplicate students to resolve when data in those fields are **accurate**.

The outcome is not to produce false TIDs (i.e. assign two or more numbers to the same student) or the converse (i.e. assigning one number to two or more different students). So clean data is an absolute necessity when it comes to assigning the initial TID (see step 11: training section) as well as establishing the right threshold that will not overproduce unique student numbers (i.e., they will create a greater burden on district staff who will need to resolve them later).

To help districts resolve potential duplicates two online applications, which merge several student records via ‘confirm probable matches’ and ‘manual match’ must be developed. This will permit districts to review specific fields to determine if the student(s) is the same (keeping only one TID), are different (keeping two or more TIDs), and merge the data elements in those selected fields that produce a correct single record on a student. Implicit in these operations is the assumption that the corresponding data fields are reliable. These applications may include 15 to 20 fields of information on each student.

One goal in the initial production of the TID is to establish business rules that tolerate a threshold of no more then 3-4% of the cases being referred back to schools for resolution (see an example of Business Rules in appendix H). With these two applications the OPI can expect to hold error creep to a minimum. This approach will also reduce burden on local school district staff to the greatest extent possible.

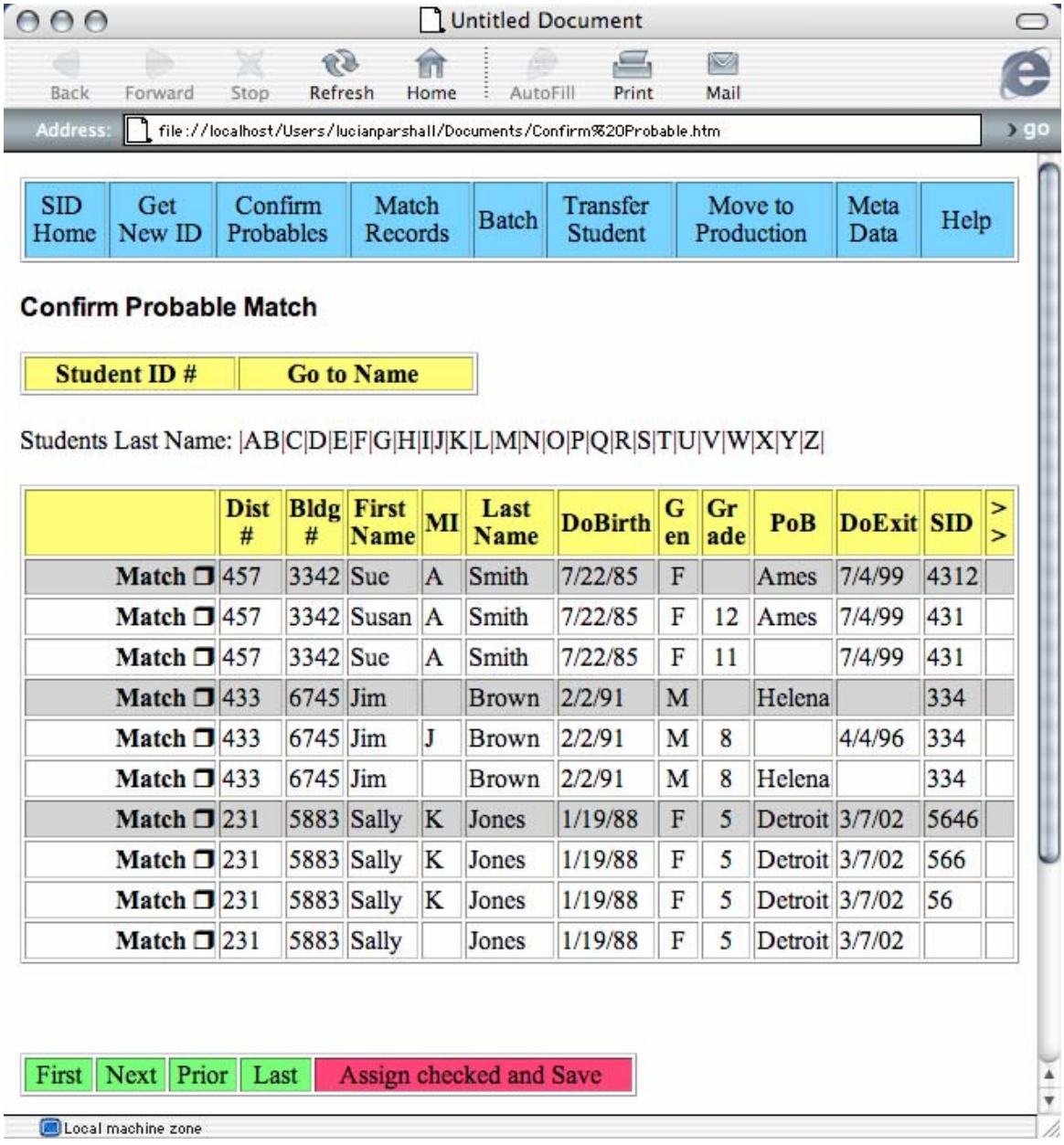
The processing of student data in these applications is based on the belief that the OPI’s responsibility is to gather and report the data from districts - not to change it. Only the source of the data can modify it. The OPI must also work under the assumption that it is better to make the error of accepting a TID that is a real duplicate than to make the mistake of rejecting a TID that is in fact *not* a duplicate. It is for this reason that about 3-4% of the students would be placed in the ‘confirm probable matches’ and ‘manual match’ applications shown in examples 1.2 and 1.3 below.

These decision rules will result in 3-4% error when comparing record-pairs (two records when examined together determine whether they refer to the same or different student). This percentage depends on the thresholds being used, which ideally would:

- minimize the number of false matches,
- minimize the number of false non-matches, and
- minimize the number of record-pairs, which must be examined manually by districts.

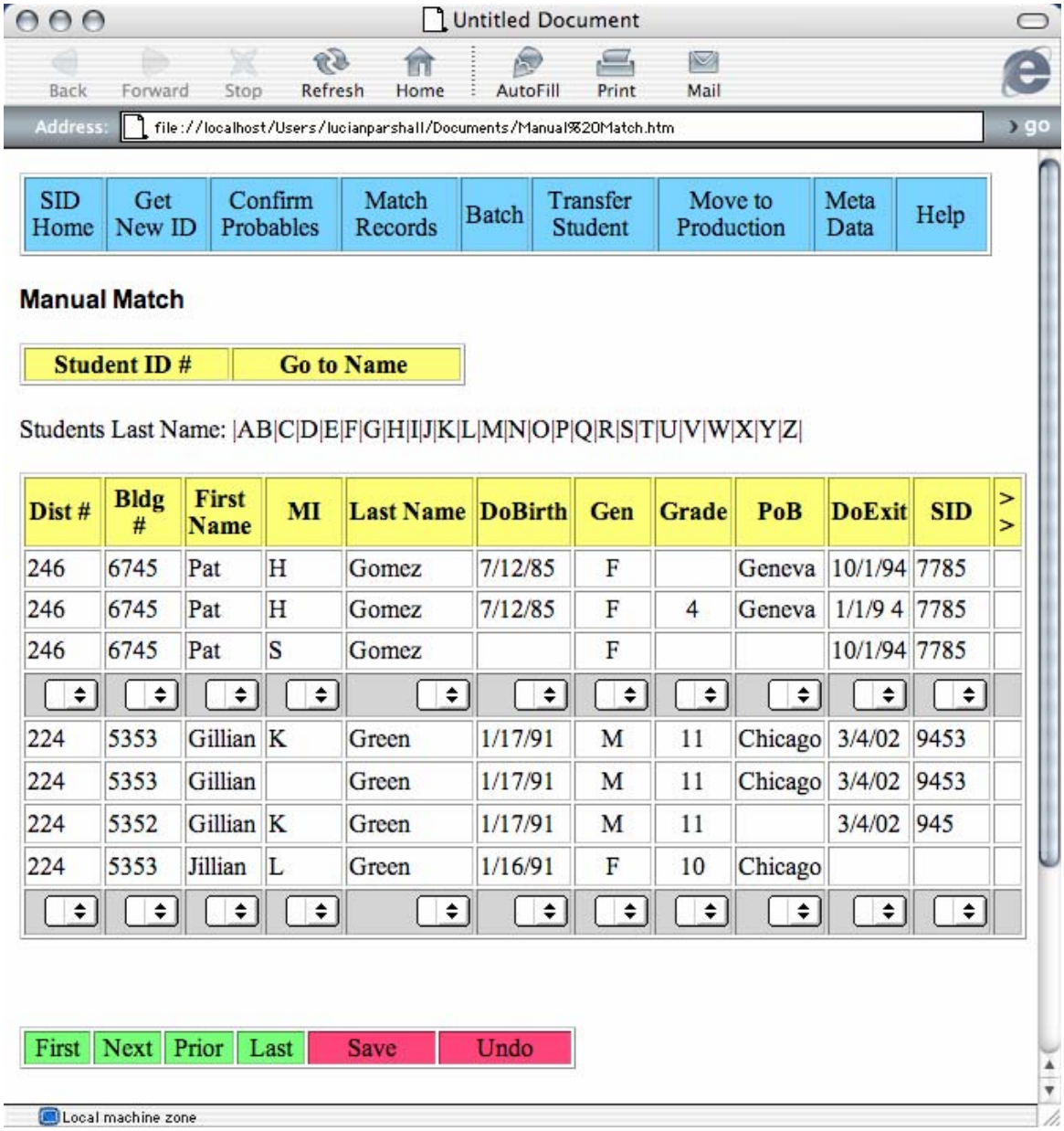
For a more graphic example of setting thresholds see appendix I.

Confirm Probable Match application example 1.2



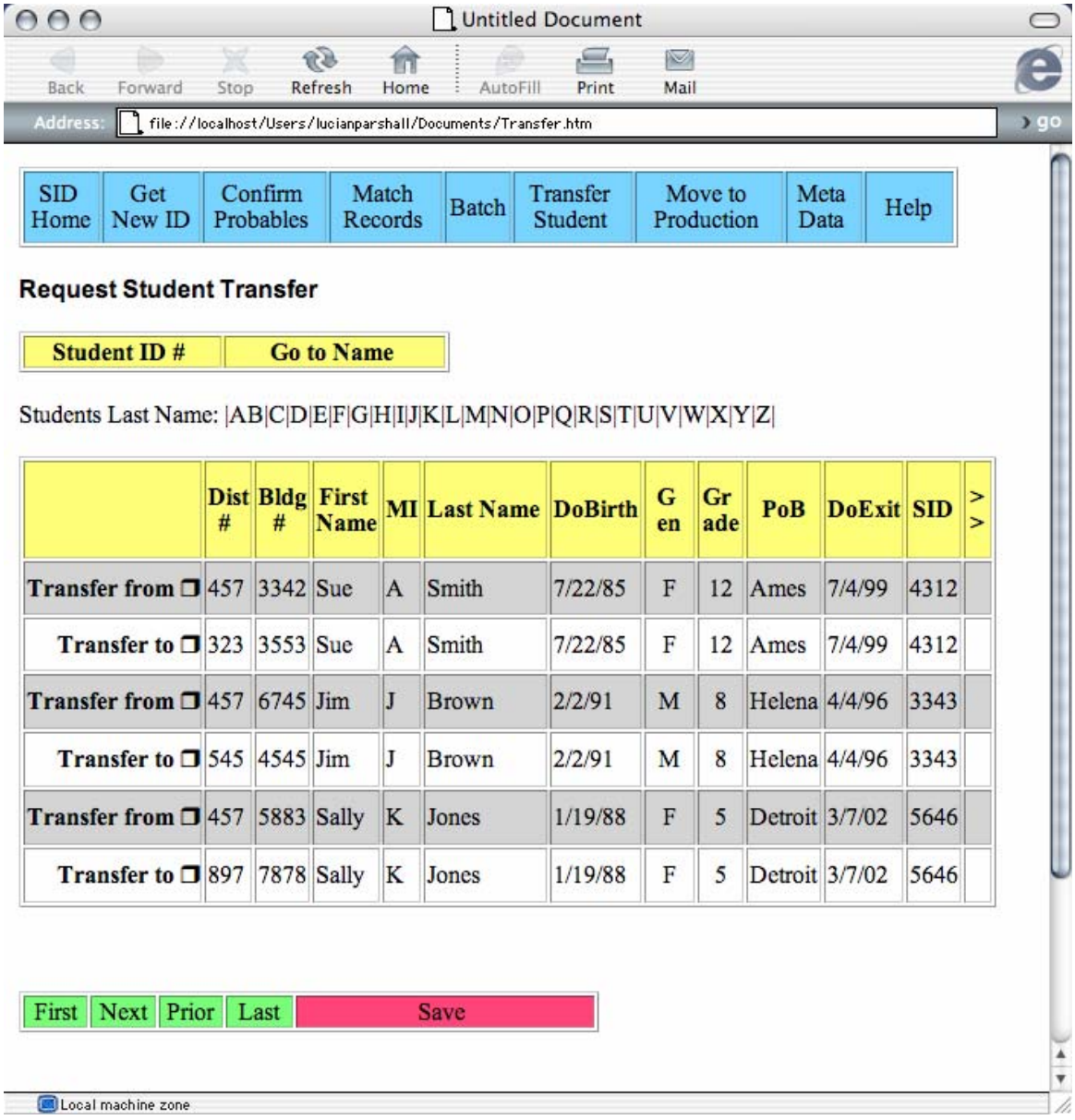
	Dist #	Bldg #	First Name	MI	Last Name	DoBirth	Gen	Grade	PoB	DoExit	SID	>>
Match ❐	457	3342	Sue	A	Smith	7/22/85	F		Ames	7/4/99	4312	
Match ❐	457	3342	Susan	A	Smith	7/22/85	F	12	Ames	7/4/99	431	
Match ❐	457	3342	Sue	A	Smith	7/22/85	F	11		7/4/99	431	
Match ❐	433	6745	Jim		Brown	2/2/91	M		Helena		334	
Match ❐	433	6745	Jim	J	Brown	2/2/91	M	8		4/4/96	334	
Match ❐	433	6745	Jim		Brown	2/2/91	M	8	Helena		334	
Match ❐	231	5883	Sally	K	Jones	1/19/88	F	5	Detroit	3/7/02	5646	
Match ❐	231	5883	Sally	K	Jones	1/19/88	F	5	Detroit	3/7/02	566	
Match ❐	231	5883	Sally	K	Jones	1/19/88	F	5	Detroit	3/7/02	56	
Match ❐	231	5883	Sally		Jones	1/19/88	F	5	Detroit	3/7/02		

Manual Match application example 1.3



A student transfer application (see example 1.4 below), which will allow the receiving district to enroll a student when the sending district enters an exit date is similar to the 'matching' application, will be developed. While not all districts may choose to use a student transfer application it does reduce error for those that do. The issue that often arises between transfer students is one of time. The sending district (Alpha) may not be aware that the student actually moved or enrolled in another district (Beta) - or worse yet in the case of a drop out have no desire or time to officially exit the student in the transfer application. Furthermore, usually the receiving district (Beta) must enroll the student immediately and may not be in a position to search a database or contact the former district (Alpha) to go online and enter the exit date.

Request Student Transfer application example 1.4



	Dist #	Bldg #	First Name	MI	Last Name	DoBirth	Gen	Grade	PoB	DoExit	SID	>>
Transfer from ❐	457	3342	Sue	A	Smith	7/22/85	F	12	Ames	7/4/99	4312	
Transfer to ❐	323	3553	Sue	A	Smith	7/22/85	F	12	Ames	7/4/99	4312	
Transfer from ❐	457	6745	Jim	J	Brown	2/2/91	M	8	Helena	4/4/96	3343	
Transfer to ❐	545	4545	Jim	J	Brown	2/2/91	M	8	Helena	4/4/96	3343	
Transfer from ❐	457	5883	Sally	K	Jones	1/19/88	F	5	Detroit	3/7/02	5646	
Transfer to ❐	897	7878	Sally	K	Jones	1/19/88	F	5	Detroit	3/7/02	5646	

In addition, it is also recommended that an 'end of cycle' match be performed by the DRA. At a specified time (usually at the end of the school year) an application will compare records in the previous cycle to the current cycle and

1. Identify unreported TIDs and unmatched TIDs that are not resolved and push into next collection cycle.
2. Remove unresolved duplicates after 24 months and forward their names to the Office of the State Auditor.

Steps 8-10 involve the creation of reports and analysis of the data
One important decision that must be made early on is to determine the number of users that will be using the special analysis tool to access the student data in MERS. These tools are commonly referred to as Business Intelligence (BI) tools.

Initially a BI tool will be used by the DRA to generate the standard reports (described in step #9) which are posted on the website for the general public. However, another purpose of this tool is to reduce dependency on the DRA to generate the unending variety of reports that are usually requested. The goal is to have the DRA provide enough training so that users can generate their own information – independently (see step #10).

Recommendations and Cost
Steps #8 through #10 address the need to purchase and use a software tool that accesses the student data. There are as many as 20 companies that sell BI tools and offer educational discounts for their products. Each can provide the OPI different purchase options.

At a minimum, the OPI would need to purchase a license for the business application that resides on the warehouse computer. This license can run as high as $10,000 per year plus a $2,000 annual renewal. Once this application is purchased there are three different approaches that can be taken.

The first option is to purchase host ports for the warehouse computer. The advantage of this is that the application does not reside on the users PC and any number of users inside the firewall can use the BI tool depending on the number of ports that were purchased. For example, if 20 ports were purchased then 20 concurrent users could access the student data at one time. The 21st user would have to wait until another user logged off. Host ports may run up to $1,000 per port plus annual updates fees.

The second option is to purchase a single stand-alone copy of the BI tool that installs on a single users PC. In this case the user has immediate access to the warehouse at any time and will not need to wait for another user to logoff. This approach is recommended for both the DRA and the DBA who will need to access the warehouse daily. A stand-alone copy can run as much as $2,000 per application plus a $500 annual upgrade fee.

The third option is providing a BI tool via the Web. In this case any number of users will have access via the Internet to query the student data on the warehouse. For this solution, special firewalls must be designed so that users can read-only approved data and not enter other parts of the warehouse that are restricted. This approach promotes the concept of an information democracy but is expensive because it allows an unlimited amount of users to access the data at any time. Web licenses may cost as much as $20,000 plus annual maintenance fees.

All the above options depends on performance (speed and capacity) of the hardware being used to house MERS data. Even a moderate number of users running short queries (less than a 5 seconds per cycle) will decrease the performance in a warehouse, thus the ability to provide a timely response to the query asked by the user. Slower performance also occurs with novice users when they submit a ‘run away’ query that has no response; they can tie up amounts of CPU time. There is nothing more frustrating to any user requesting his or her first query than having to wait 4-5 minutes for a response.

Step 8: Select a Business Intelligence (BI) Tool

Step #8 involves the selection and purchase of a Business Intelligence (BI) tool. While these tools are not needed initially by a large number of end users other than the DRA, they become critical as the number of users increase, as the educational community begins to ask more complex questions and as the size of the warehouse grows.

Often end users will use a decision support (a.k.a., business intelligence) tool out of impatience with the IT department. Or, the IT department gives the user these tools to relieve the pressure off of itself. Selecting a BI tool is based on the belief that stakeholders in the educational community can be more supportive if they have a better feel for what the tools are actually used for. In doing this many of the 'big brother' fears are frequently reduced. And selecting the right BI tool can make picking that virtual needle out of that virtual haystack a lot simpler.

One of the main uses of a decision support or BI tool is to check that 'everything' is okay. The DRA in many cases often write reports that could hardly be called analyses (i.e., to provide a report 'of record'). Not much will be done with many of the queries and reports created with BI tools. They are run to confirm a staff person's fuzzy but intuitively felt notion of 'okayness' (i.e., to confirm the 'obvious'). BI tools do not often reveal anything amazing that a user doesn't already suspect. But the information produced with the tools gives them confidence in the data and help them figure out how something 'works'.

Most users in the educational community who use the decision support tool will not be looking for some grand unified theory of how firm XYZ works. Rather, they want to understand some small aspect of a student, like what are some of the variables that makes student A always be on time while student C be tardy.

Sometimes BI tools can be used in a side-by-side comparison of a series of measures. Sometimes this is identification of the most, the least, the earliest, the latest (i.e., to compare the same type of information in different time periods) or to check student performance versus formal and informal constraints/resources (i.e., to identify the out of the ordinary).

Usually the user has some vague criteria of what is out of the ordinary. A decision support tool does kind of double duty in that it helps refine the criteria of what is out of the ordinary and helps identify what new data is needed to fit the refined criteria. For all kinds of reasons it is often necessary for the educational community to agree, "these are the final numbers". Often not everyone agrees on all the data but just enough data whose credibility must be accepted for actions to be taken. Decision support tools often are used to produce this "official" information (i.e., to confirm and sometimes to discover trends and relationships).

BI tools provide value to states and districts in several ways. But most importantly, these tools help short-staffed SEAs by facilitating rapid, sophisticated analysis of thousands of data elements to sequence, cluster, and frame. The DRA can then use the tools to present information to other stakeholders in a way that is more easily understood. This is because the hundreds of tables in a data warehouse have extensive relationships with each other that cannot easily understood by users accessing the data. It is for this reason that an interface (a switchboard) is often created by the DRA to make the warehouse friendlier to the user (see examples of how a complex warehouse model can be reduced through a 'switchboard' to be accessed with a BI tool in appendix J). By providing easy but secure access to a relational warehouse with more powerful BI tools, two long-standing problems— dynamic and static reporting— can be addressed. BI tools are more capable of handling queries and are more complex than formulas in standard database or spreadsheet software.

Step 9: Generating Reports

Steps #9 and #10 also involve both the 'push' of information and 'pull' of data as described in the *information democracy* above. Step #9 involves preparing and posting standard state and federal reports.

The State Report Card (Sec. 1111(h)) of NCLB requires graduation rates for public secondary school students, comparison between achievement levels as well as other optional items such as: attendance rates, average class size in each grade, the incidence of school violence, drug abuse, alcohol abuse, student suspensions, and student expulsions, the percentage of students completing advanced placement courses, and at least one other academic indicator, as determined by the State for all public elementary school students.

Both the State Report Card as well as the state Adequate Yearly Progress (AYP) report must disaggregate student data by: race, ethnicity, gender, disability status, migrant status, English proficiency, and status as economically disadvantaged, except that such disaggregation shall not be required in a case in which the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student. Finally, the annual state report on high school graduation and dropouts will be much richer and contain greater detail (see Reporting Exit Rates in appendix K).

DRAs can save time and resources by setting up report writers to run reports periodically and by running and presenting queries that aggregate and disaggregate data based upon user-defined criteria. Districts can access the DW and run queries in a matter of minutes that would have required months of analysis without a data warehouse and an integrated report writer (part of the BI tool kit). An integrated BI report tool also minimizes the need for states and districts to purchase complex reporting solutions that require upfront and ongoing integration services and staff training.

In this respect, the OPI will be able to create routine static reports when responding to requests and post them on a Web page for the general public. Analysts will have the ability to create more complex reports upon request. The DRA may also be able to group related reports, so consumers can perform limited drill from one report to another (i.e., using a data mart). This structured but limited control of creating and deploying standard reports and posting them on the Web provides a managed environment that lets consumers easily access as much of the information that is relevant to them, when they need it.

The MERS model involves the use of the OPI Web page to report the data in a format close to the familiar 'paper form' that the educational community is used to viewing. Once the MERS is operational users are able to create dynamic reports by querying (the process of extracting data from a database and presenting it for use) the warehouse and quickly assembling the specific fields of information that are needed. This is usually done online, but selected student data files can be off-loaded to reside on PCs or client servers. Because more powerful analysis tools can present relational warehouse data in an intuitive context, data users are better able to mine the information they wish to query. With more advanced software, queries can be created 'invisibly', and users require no knowledge of a low level program language or a high level statistical package (SPSS, SAS) to retrieve desired results.

Step 10: Data Mining

In step #10 a user has the ability to access the student data to extract unexpected patterns. Data mining uses pattern-recognition (found in most BI tools) to obtain different views and new perspectives on existing data. Data mining is about ratios, patterns, and influences on decisions. After all, data are one of the few educational resources that are completely reusable.

Decision support tools are often used to address problems highlighted by the traditional statistical packages (SPSS, SAS). This is because decision support tools can be aligned with communications to provide real-time alerts, or automatic actions for administrators, teachers, and staff.

States that utilize decision support tools foster an environment of continuous improvement. Decision support tools create immediacy around important educational issues, and promote proactive responses by policy makers to school improvement issues as they arise. This sense of proactive responses enables states and districts to employ a problem-solving approach to issues as they surface.

Not only do data warehouse users want to aggregate data in their own terms; they may also want to disaggregate them in their own terms. For example, the user may want to show total dropouts for a given year. The user may then want to be able to click on the data and have them explode into dropouts by month; to click again and have the data explode into race/ethnicity by month or by age. BI tools must meet this need for a specific set of drill-down requirements that vary by user and by task. In fact, sometimes the user does not know how they want to drill down until they see the data and start drilling.

Traditionally, the utilization of the information that is hidden in the data has been achieved through requesting specific program routines or conventional statistical packages, such as SPSS or SAS. The traditional approach involves a user formatting a theory about a potential relationship in the population and converting that hypothesis into a query. This is a somewhat straightforward, top-down approach to data analysis. The difference with data mining is the online interrogation of the student data, particularly when it is in a relationship with performance scores and school infrastructure data sets. In other words, data mining tends to be a self-organizing, bottom-up approach to data analysis, whereas traditional statistical approaches are more verification-driven.

Many researchers warn against data mining in the search for new results. Such a warning is accurate when applied to traditional statistical approaches. When several statistical tests are performed, one drastically increases the risk of ‘finding’ something that really does not exist. When users test for patterns of statistical significance because they suspect patterns will illustrate cause and effect, then the users are likely to be fooled into finding what isn’t there. In short, warehouse users will have the ability to perform more powerful analysis but also have a greater opportunity to misinterpret and present misinformation. Understanding what information can be gleaned from the data often takes a seasoned evaluator.

Moreover, the ability to mine data promotes a different attitude and process for data users—an attitude that engages curiosity about regularity, pattern, exception, and a process that ‘peels the onion,’ so that one will look beneath the obvious for more subtle (and often more interesting) patterns. This approach makes few assumptions about the data and deliberately accommodates the unexpected.

Step 11: Conduct Training

As a rule of thumb, the closer the data gets to the classroom, the more professional development services will need to make available to instructional staff. States and districts that wish to use data effectively to drive decisions must not only have the right technologies in place, but must also have trained their users to take full advantage of the DW.

Step #11 recognizes that:

- change is a process, not a single event and as such it must be facilitated by a multi-dimensional effort;
- as the user moves from simple awareness to actual integration of the student data set, ongoing support is vital;
- support must ultimately come from local resources where they are readily accessible to users;
- as users move towards adoption of the MERS they will proceed through the predictable stages that require long-term training and support. By OPI.

This approach to training can best be summarized in the belief that a well-established support system is critical before MERS can be integrated into day-to-day decision-making.

However, not all professionals are at the same degree of concern nor the same level of understanding of how change occurs. It is because the base of users of the student data will continually increase that there will constantly be varying levels in their understanding of what the MERS can provide.

The ultimate goal of training is to have everyone who touches the data at every level know what is expected of them, so that the data that are submitted will be the most accurate. To achieve this goal four levels of training are necessary to:

1. Help the educational community understand policy, funding, potential uses of the student data, MERS, Meta data, and confidentiality issues;
2. Assist vendors in: the inclusion of key fields that may be added to their SIS, exporting data to OPI, importing the TID back into their SIS and understanding error traps being used by the OPI;
3. Provide hands on instruction to key district staff on how to use the online applications that manage a student record as well as the TID; and
4. Provide users with an overview of BI tools and how they can be used for ad-hoc queries.

Training must occur at three levels. First, awareness sessions are necessary that relate to the Meta data. The educational community must be aware of the fields of data that are being requested, when they must be submitted, how they will be edited prior to submission and what variables are expected within each field so that the data can be gathered by a local district prior to being sent to the OPI.

One of the critical parts of the Meta data awareness session is what and how error traps work on each student field (see taking Data to Laundry in appendix L). The results of an edit check on a student data set (i.e., the list all of the student records that are not valid and need correction along with an indication of the inconsistencies that were identified) must be described to the target audience.

Awareness of Meta data can be delivered in three or four large group sessions across the state. Sessions usually take a full day and while a variety of staff are welcome, they target key

individuals in each district that handle student data. Usually a separate all day session is necessary for vendors of existing SIS packages in Montana to inform them of the edits that will need to be incorporated into their student management packages.

The second training relates to use of the online applications. It is anticipated that the focus of this training would occur in two areas:

1. Training staff in regions and larger districts (e.g., Helena, Billings, Great Falls, Missoula, etc) on the various student entry processes and application functions (i.e., state-wide TID incorporation back into the local district SIS, batch processing, data submission timelines, duplicate student resolution, matching and transferring).
2. Training of the vendors on the various file submission routines, import processes, data formats and export processes to incorporate the state issued TID into their SIS. Vendors also include those districts that have developed their own 'home grown' student management systems.

The key point to this level of training is that participants embrace procedures that deal with the complexity of editing data beyond technical orientations often provided on how data collection applications function. Training must combine technology with an orientation of how education operates at the LEA, SEA, and Federal levels, to provide meaning to users across the educational community, so that users will trust the data and apply data results when making decisions. Training must also include processing the following:

- Who gets notified when an error is discovered and how is notification done?
- What is the procedure for making corrections of data within an agency (i.e., who actually makes them and retransmits the error-free data)?
- Who reviews, verifies or signs off on the cleaned data?
- Who provides technical assistance to the end user?
- What is the procedure to ensure a new copy of the data is retained for auditing?
- Who receives confirmation that the file has been received as specified?
- Who secures the data and maintains confidentiality?

It is estimated that up to six to eight specialists would be needed to conduct these workshops with users in each region. These training sessions are anticipated to occur just before the actual application goes live to ensure as accurate data as possible once the system is active. These workshops would include OPI staff.

Finally, users must be provided with an overview of BI tools and how they can be used for ad-hoc queries. This awareness training can be delivered in three or four large group sessions across the state with test data (usually the first student collection is considered test data). Sessions usually take a full day and while a variety of staff are welcome, they target key individuals in each district that handle district report cards, AYP and other local issues such as attendance and drop outs.

Phone/email support would be available after training. Online technical support in the form of help menus will always be available.

The need to train potentially over 500 key individuals in Montana's 459 school districts on the creation and submission of clean and accurate data is essential to the successful deployment of the TID. Without accurate data at the front-end that is used consistently across the state, the ability to perform consistent and accurate comparative analysis and reporting will be limited and frustrating to the educational community.

Recommendations and Cost
One-day regional workshops in three to four regions would consist of 3 trainers per workshop at an honorarium of $350 per day. These awareness workshops would cover such items as the overview of MERS, Meta data, reporting, security, and OPI collection time tables. Trainers would be selected from staff who participate in the data managers committee. Total cost for three to four regional workshops would range from $3,150 to $4,200.

Specialists who deliver the training relating to the online student Web applications will be used. The initial hands-on sessions will be six hours. Follow-up sessions may be necessary. The number of specialists used for this training is determined by the number of sessions needed to train all the staff in Montana districts that touch the student data. Two specialists per session would be used and limiting the number of participants to 25 per session, and assuming that approximately 500 local staff would need to be involved in hands-on training, 20 sessions will be necessary, with a possibility of 5-6 follow-up sessions. Specialists are paid a rate of $30.00 per hour, plus mileage. This brings the hands-on training expense to approximately $460 per session. Total cost for three to four regional workshops would range from $9,200 to $11,960.

The project will need to print documents used in the training for staff to refer to when they return to their local district and use the online applications. Materials will be limited to approximately $12 per person. Total cost for printing for workshops and hands-on sessions would be $8,000.

The total cost for training (per year) is $20,300 to $24,160 in 2004, assuming that OPI will secure computer labs in local districts for the 20 to 26 workshops at no cost and that OPI will also secure rooms for the three to four large sessions at no cost.

Step 12: Supporting Decision-making

Step #12 involves decision support. Data warehouses exist to support management decision-making.

One defining characteristic of data warehousing is the separation of transactional systems from decision support systems (DSS). By separating these two very different functions (see chart below), the data warehouse architecture enables both operation and decision support applications to focus on their distinct strengths and provide better but different performance for each of their end users.

Transaction Systems	Decision Support Systems
Day to day operations	Historical
Real time	Points in time
Updates daily/weekly	Updates monthly/quarterly
7x24	6x18
Read/write	Read only
Short term data retention	Trends, patterns or longitudinal data
Mission critical queries	Strategic or analytical queries
More open access paths	More restricted access
Standardized reports	Ad hoc reports
Server based	Warehouse

When building a transactional system, the more users one talks to, the clearer the function of the system becomes. Operation systems are examples of transactional systems, such as grants' payments or purchasing processes; they generally update a record, a single event, one account, an

inventory item, or one order. Transactions are generally predefined and require the application to provide access to one record at a time.

When designing a decision support model, the opposite is true. The more users one talks to, the greater the variety and scope of the system. It is up to the Data Resource Administrator to find out what data users want to address and then build the system to deliver that information. The nature of a DSS is one of discovery. If one walks up and asks a data user what s/he wants, the user usually cannot articulate what it is. But, if one shows a data user the possibilities are, then the user starts to burst forth with all kinds of requests. The DSS user begins to operate in a mode of discovery, typified by the mind-set: "Aha! Now I see what the possibilities are!"

With the enactment of NCLB legislation over 33% of the states have initiated a DSS and another 57% are in the design or discussion phase (ibid, 2002). However, support systems in education are in their formative years, primarily because:

1. Decisions made by school administrators, teachers, and boards of education are not well modeled. Policy formation is often poorly structured, inconsistent, and highly personal (i.e., not subject to a flow chart).
2. Educators must acquire and maintain new skills beyond curriculum and instruction; this includes technical, analytical and reporting skills, and
3. The warehouse technologies upon which DSS are built require ongoing funds, priority, and maintenance—none of which are easy to project or manage.

One soon realizes that achieving a high level of decision support by many of the users takes time to deploy. Any agency will soon recognize that users of the warehouse progress through stages. Recognizing these stages can help organizations move through them more smoothly towards a dynamic DSS system. Such stages can be categorized as follows:

1. Individual Impact: The first stage consists mainly of individuals with the greatest need (usually those who will derive immediate benefit) to produce familiar reports for boards of education or policy makers. This is a somewhat mechanical use of the data in the warehouse and often means the reproduction of reports in the former aggregate formats.
2. District Success: In the second stage, the OPI maximizes impact of the data by deploying more powerful tools that produce more substantial benefits to school districts. These tools lead to clean, reliable data, improved analysis, and customer satisfaction. In this way, local districts gain greater insight from the higher value found in the information that they placed in the warehouse. District reports are often more focused on student performance, disseminated to a broader group of stakeholders and in some cases may be more comprehensive (i.e., include variables that impact performance beyond assessment results).
3. Educational Community: In the third stage, the warehouse becomes the single source, which is shared completely by all stakeholders. As a result, everyone is better informed about significant issues that can affect student performance, and ultimately prepare students for adult life roles. With the facts at hand, stakeholders are better able to promote decisions that align with common practice, policy, or legislation. At this point, the educational community becomes truly coordinated. At these stage 'ad-hoc' queries rather then summary reports are more common. The OPI datasets that have the highest use are farmed out into data marts.

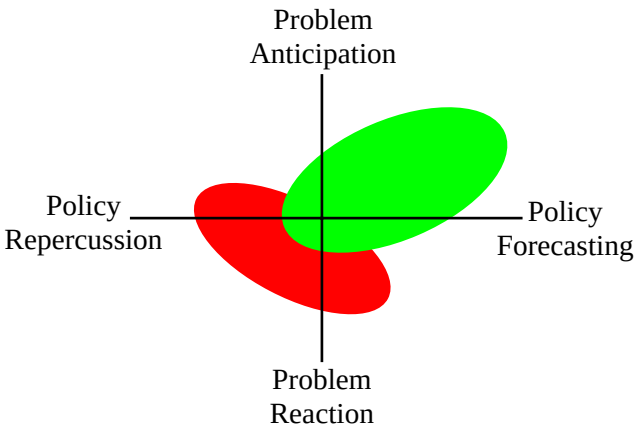
As the DW evolves into a DSS it must stimulate the end user's imagination. The first iteration explicitly describes to the data user exactly what some of the possibilities are for analysis and creates the need to populate the warehouse with reliable data as quickly as possible. Until that first phase of the model is operational, the DSS end user simply does not have the experience to declare how great a value it might be. Often the first iteration has a crystallizing effect on the

imagination of the data user, but does not produce profound results because certain data sets may not have matured.

The MERS by its very nature requires that change be an inherent part of the system. While requests for structural changes to the warehouse can be expected, it is assumed that the majority of the structural changes will require reconfiguring some of the elements that already exist in the fields, rather than enlarging the size of the record layout.

The fact that there are changes in the DW over time is a sign that it is gaining maturity. To make many changes from the first to the second iteration is a sign of progress and health. Usually, by the third iteration adjustments tend to be small. This is because the data user goes through the learning and discovery process early on.

As the warehouse continues to accumulate and store ever larger amounts of data, it is archiving knowledge which can help the educational community carry out better policy forecasting and problem anticipation (green oval below), rather then constantly reacting to problems and responding to outdated policy (red oval below).



The use of data in the warehouse will be a journey of discovery. At each step one tries to learn more about the data, refine the model, and form new questions from the derived results. In the end, the exploratory analysis is likely to generate worthwhile questions that may go beyond the data at hand. This will guide future research and produce more informed decisions. Using the MERS warehouse, data analysis becomes part of the larger process of creating better educational policy.

When the DW is fully operational the educational community becomes aware of many challenges; however, the need to prioritize them, delegate responsibilities, help create effective solutions, and measure the success of these solutions become part of the ongoing strategic plan. Administrators who have led such initiatives stress that the most challenging part of the process is transforming the way in which colleagues, parents, and staff view and use decision support tools to inform the community about better policy. For a complete list of answers that can be generated by the MERS when it is fully operational see appendix M.

Recommendations and Cost

Decision support systems are as fluid as the changing needs of the decision makers' requests for information. Different people want different information, informational needs naturally change over time, and supplying information generates the demand for more information.

With MERS, while some data sets are maturing, other sets will be starting their cycle toward maturity. At any one time, one data set (e.g., student performance data) may be well into its third iteration of development, and another set (e.g., quality data on professionals) may be undergoing its first. Given the nature of data warehouse development (where one phase is built and then another), in all likelihood the users of the DW will exist in different stages of maturity.

It is for this reason that a cost cannot be assigned to this step. However, past experience shows that as the DRA assists the initial users of MERS in the use of the BI tools, in turn they assist the novice users on how to manipulate the student data to achieve the outcomes of the queries being requested.

Project Plan

The OPI must establish two key user committees to support and participate in the construction of MERS. These are a:

- Data Management Committee; and a
- Project Steering Committee.

These two committees could be comprised of the overlapping individuals. The committee responsibilities are summarized as follows.

Data Management Committee

The Data Management Committee will provide focused OPI project leadership and scope management for the duration of the 12 steps with emphasis on setting the scope for each phase within the time and cost boundaries for the project. This committee is chaired by the DRA and consists of the DBA, technical staff from local districts and data entry staff. In effect they provide oversight of the project management plan (see appendix N for a draft project plan). This includes: mediation with test sites, decision making, change control review as well as advice relating to procedures requiring OPI approval, issues resolution, and final acceptance of deliverables from each phase. This committee meets monthly. Responsibilities may include:

- Expedite the resolution of issues and/or actions impacting the on-time review and approval of the project plan and phase scope,
- Expedite the resolution of issues and/or actions impacting the review and approval of the change control process,
- Expedite the review and required feedback for reports, queries and deliverables that will require OPI approval,
- Expedite the resolution of any issue, dependency and/or outstanding action impacting the on-time completion of each phase of the project,
- Monitor the OPI portion of the change control process,
- Monitor the review and approval of the biweekly project status reports,
- Provide support to and feedback to the Project Steering Committee, as may be required from time to time.

Change during the life of a project is a reality that cannot be avoided. For a project to be successful though, change must be carefully managed. A seemingly minor change can have dramatic consequential effect on the project if it is not managed properly. The change request process will provide a formal means to:

- Request a change
- Identify the impact of a change
- Confirm that the appropriate parties review and approve the change
- Manage the change
- Maintain a record of the change

The project work plan must be prepared and will be updated regularly to allow comparison of progress to plan. Reporting will include, but not be limited to, the following:

- Status reporting— Activities that are behind schedule, progress vs. plan, project issues that

cannot be resolved independently by the project manager,
- Progress reporting— Activities accomplished during the period,
- Forecasting— Activities scheduled for the following period,
- Status meetings— During the monthly meetings conduct reviews to verify that commitments are being met, that issues are resolved in a timely manner, and that quality standards are maintained.

Project Steering Committee
The Project Steering Committee will oversee warehouse activities required to satisfy OPI responsibilities, including facilitation of all collections, results of pilot site testing, data element revisions, review of the data model (for understanding), and formulation of plans for data quality improvement. This committee is chaired by the Assistant Supt of OPI, and consists of key central office administrators from local districts, representatives from key statewide organizations, OPI staff from each program area and from time to time representatives from the Data Management Committee. This committee meets quarterly.
Responsibilities may include:
- Oversight and participation in data-related activities from addressing district level policy concerns, modifying fields to the Meta Data, targeting the phase in of collection dates, assessing data quality issues and approval as well as format of statewide reports,
- Make recommendations to the OPI regarding the approval or rejection of deliverables, data quality improvement recommendations and resolution of interagency related issues.

RFP Deliverables

Deliverables:
1. Develop an application (with clear business rules) that generate unique TIDs with the following characteristics:
- A number that is only assigned to one student (is unique).
- A number and name that can be confirmed as being correct (is verified via check sum).
- A number that meets all criteria as an identifier (is valid).
- A number that has no intrinsic meaning (is nominal).
- A number that can be substituted for a student's name (is not personally identifiable).
- A number that is permanent over the lifecycle of the student (0-21 for special education).
- A unique number that is returned and used by all local districts (is ubiquitous).
- A number that is issued only by the OPI (is restricted).
- A number that is accessible by selected OPI employees only (is confidential).

2. Four Web interface applications for districts
Assuming that the assignment of the TID will produce approximately 3-4% of the students as duplicates, two applications are needed to prevent error creep. One to 'confirm probable matches' and a second to 'match records' of the potential student(s). A batch application (ETL) must be developed to transfer large sets of data, consisting of an online edit check and a correction tool for the user/district. Finally, a transfer application, is needed to allow the receiving district and the sending district to move student records. All Web applications my have an 'administrators' application that permits the change of high-level data elements or definitions by OPI staff without the need for recoding by a programmer.

3. Audit application for MERS
An audit application that will place internal controls within the above applications that ensures that when a student record is accessed within the master file or during a transfer between districts, a historical log is kept of the transaction.

4. Meta Data Dictionary
The incorporation of an online Meta data manual with the agreed upon data elements that will provide a description of and the meaning to the values in each field, so that the user can understand what data to manipulate, where it can be found and how it should be exported.

5. Error Trapping Rules
A list of edits and business rules that are used in the online applications, also provided to the vendors of student information systems for their use with local districts. These are authorized and issued by the OPI.

6. Training
Training includes awareness sessions, training workshops, and ongoing technical support that will focus on the entry of student data at the school and district level. The need to train key data entry individuals in Montana's school districts on the creation and submission of clean and accurate data is essential to the successful deployment of the MERS. Without accurate data at the front-end that is used consistently across the state, the ability to perform consistent and accurate comparative analysis and reporting will be limited and frustrating to the educational community.

7. Technical Documentation
All aspects of the MERS as well as the Web applications must contain sufficient documentation for a new staff person to maintain and modify the application in a reasonable time. Documentation within the applications code is sufficient. In addition, there must be administrative documentation (for OPI staff to make hi level ongoing changes to an application or modify standard reports) in both hard and soft copy.

8. Recommending a Business Intelligence tool
One of the key elements to data warehousing is that it helps the community move beyond how much data to collect and how often, onto issues that focus on timely reporting and evaluation. The MERS warehouse will prompt the educational community to ask more detailed questions about their concerns. For example:

- The general public may use data to become more knowledgeable about educational resources or to make informed decisions about schools of choice.
- The media may use data to inform the public about such matters as student performance or program expenditures.
- Educational associations may use data for lobbying and membership support.
- Institutions of higher education may use data for estimating freshmen enrollment, district graduation ratios and to compare student aptitudes.
- State and local governments may request data to study education problem areas, to project funding, and to pinpoint regions in the state with growing or declining enrollments.
- Researchers may perform specialized analyses to gain a better understanding of complex relationships, trends or correlations between learning and the environment.
- Businesses may use data to forecast demands for their products, predict supply of future employees or to sell homes in desirable school districts.

The RFP must provide a list of the strengths and weaknesses of the several BI tools for selection and use by OPI and the educational community.

8. Recommending the Right Technology
The recent explosion of network servers has its place in any computing platform, but in general, today's server farms are not designed to handle extremely large data sets, widespread and sizeable numbers of concurrent users, or extensive throughput (the combination of speed and capacity to

move data). Organizations move to warehouse computers for three main reasons: cost effectiveness, performance, and economy of scale (potential to support growth and change).

Cost effectiveness: Typically, large-scale warehouse platforms have superior price per performance compared to more traditional mainframes. Traditional architectures achieved their moderate performance levels by using expensive proprietary microchips and finely tuned co-processors. These components are often expensive to design and expensive to fabricate. With warehouse computers, the philosophy is very different. The idea is to use mass produced and inexpensive 'off the shelf' components to achieve high performance. These standard, mass-produced chips and processors are far less costly—the economics of mass production are in one's favor.

Performance: Parallel processing is the key to this high level of performance. Dividing a larger problem into smaller pieces, and coordinating the execution of those smaller pieces, is the fundamental concepts behind parallel processing. For example, traditional mainframes have one large single processor that does all the calculations. Today, warehouses have a pool of processors. By dividing the workload, they achieve a much higher performance level. They also achieve a high input/output performance by spreading the data across a large number of disks (called 'striping'), thereby distributing the input/output requirements and increasing the access speed to the data located on the disk.

Scalability: Warehouses are able to expand and accommodate the rapidly growing and changing needs of the educational community. Correctly designed data warehouses keep pace with the organization's needs. This yields two important results. First, an organization will be able to employ new problem solving techniques, which will build a solution that keeps pace with an expanding problem before the problem outgrows the solution. Second, one greatly increases the return on investment, because an application remains useful far longer. Building a successful data warehouse requires not just scalable technologies, but also scalable applications. This means that various data sets in the warehouse can be easily distributed across those who supply the data. Adding another relation to the functionality of the data mix becomes a natural process.

Data warehouses typically increase in size by 500% in the first three years of operation.